

Reverse Engineering in de Praktijk:

van P + R + O + D + U + C + T naar PRODUCTQUALITY

ir. T.P.H. Troquay

Agrimathica

Postbus 848, 3900 AV Veenendaal

ir. C. van Diepen

Staring Centrum-DLO

Postbus 125, 6700 AC Wageningen

Inleiding

Het Staring Centrum-DLO (SC-DLO) en Agrimathica hebben recent een omvangrijk project voor de EG uitgevoerd. Het betreft een systeem voor het Joint Research Centre van de EG in Ispra (Italië) voor oogstramingen op Europees niveau. Agrimathica is in een later stadium bij dit project betrokken, speciaal voor het engineering-gedeelte. Het project bevatte enkele organisatorisch en technisch zeer interessante aspecten, waaruit de nodige lering getrokken is.

Na de eerste fase van het project is overgegaan op re-engineering. Gebleken is dat (re-)engineering niet alleen methodische en technische aspecten omvat. Het blijkt dat de uiteindelijke kwaliteit van software wordt bepaald door een groot aantal onderliggende factoren. De letters van PRODUCT geven de sleutel voor de kwaliteit:

- Professionalism,
- Resource Planning,
- Organization,
- Development Methodology
- User Involvement,
- Coding Practice / CASE / Conventions, and
- Tools and Technology.

In dit artikel wordt ingegaan op het uitvoeren van het project, met name de inzet van 4GL en CASE voor technisch-wetenschappelijke toepassingen. Tevens wordt aandacht geschonken aan de overige rand-

voorwaarden om tot een goede systeemontwikkeling te komen.

CGMS

CGMS staat voor Crop Growth Monitoring System. De Europese Gemeenschap heeft voor het uitvoeren van haar beleid behoefte aan zo actueel mogelijke informatie over de te verwachten landbouwproductie in het lopende seizoen. SC-DLO ontwikkelde, in opdracht van het Joint Research Centre (JRC) van de Europese Gemeenschap, een systeem dat de oogstvoorspellingen over het lopende landbouwseizoen maakt. Het systeem is gebaseerd op het gewasgroei-model WOFOST, database- en GIS-technologie.

Het systeem bestaat uit drie deelsystemen:

- 1 *weersysteem met kwalitatieve informatie*; om aan te geven of het lopende seizoen warmer, kouder, natter of droger is dan gemiddeld. Dit is 'monitoren' van het weer;
- 2 *gewassysteem met kwalitatieve informatie*; om van een bepaald gewas aan te geven in hoeverre groei en ontwikkeling afwijken van het normale patroon, voor zover dat met het WOFOST-model is na te gaan. Dit is 'monitoren' van het gewas;
- 3 *gewassysteem met kwantitatieve informatie*; voorspellingen om de gesimuleerde opbrengst te vertalen in een oogstraming in kilo per hectare.

In deelsysteem 1 zijn weergegevens voorhanden van ca. 700 meetstations, verspreid over Europa, met dagelijkse meetgegevens vanaf ca. 1975. De omvang van deze database is ca. 350 Mb. Het systeem bevat software om op basis hiervan zgn. gridweer te berekenen. Hiervoor is de EG ingedeeld in ca. 1300 gridcellen van 50 x 50 km. Voor elke cel berekent het systeem het weer via specifieke interpolatie-algoritmen. Dit weer kan worden gepresenteerd op kaarten, waaruit duidelijk blijkt of bepaalde gebieden in Europa dusdanig te lijden hebben van extreme weersomstandigheden, dat oogstproblemen of -overschotten te verwachten zijn. Dit gedeelte van het systeem omvat software om de Oracle-database te onderhouden, routines om gridweer te berekenen en om de informatie te visualiseren. Dit laatste gebeurt met GIS-software gebaseerd op Arc/Info.

Met deelsysteem 2 is het mogelijk om gewasgroei te simuleren. Hiervoor wordt een tijdstap gehanteerd van één dag en de resultaten worden per decade (10-daagse periode) opgeslagen. De input voor de simulatie wordt gevormd door het gridweer, gewas specifieke groeiparameters en de bodemeigenschappen. De gridcellen zijn daartoe ingedeeld naar bodemtypen volgens de bodemkaart van de EG. De bodemgegevens zijn afkomstig uit een Arc/Info-database, waarin van elke gridcel de bodemsamenstelling is aangegeven. Hieruit kan de bodemgeschiktheid voor elk gewas worden afgeleid. Gecombineerd met het weer en de gewaseigenschappen kan het

groeiverloop voor elk gewas worden afgeleid.

De gesimuleerde gewasgegevens worden in de database opgeslagen. Ook hierbij kunnen met de visualisatiemodules uit Arc/Info kaarten worden gemaakt. Afhankelijk van de keuze kan informatie op land, regio-, gridcel- of een nog lager aggregatieniveau worden gepresenteerd. Dit gedeelte van het systeem is het meest complex en omvangrijk. Records met gewasgroeigegevens voor 15 gewassen, over 15 jaar met 36 decades met ca. 2500 calculatie-units; de betreffende database bevat daarmee enkele miljoenen records!

Het hiervoor besproken deelsysteem 3 voorziet in gesimuleerde gewasinformatie, die kwalitatief van karakter is. Gegevens van verschillende gebieden kunnen met elkaar of ten opzichte van voorgaande jaren worden vergeleken. Deelsysteem 3 omvat het kwantitatieve systeem dat gericht is op uiteindelijke oogstraming. De gesimuleerde waarden worden daartoe vergeleken met de daadwerkelijke historische productiecijfers uit de regionale landbouwstatistiek van Eurostat. Dit levert regressie-coëfficiënten per gewas en per regio, waarmee de gesimuleerde actuele opbrengstwaarden gecorrigeerd kunnen worden. Samen met gegevens omtrent beeelde arealen kunnen hieruit de opbrengsten worden voorspeld. Een globaal beeld van de systeemonderdelen wordt gegeven in figuur 1. Een voorbeeld van de modeluitvoer is te vinden in artikel Bregt en Lentjes (AI jrg 6, nr. 2, april 1993).

De eerste ontwikkelfase en gebruikte tools

In eerste instantie is de systeemontwikkeling gestart voor deelsysteem 1 en 2 en werd uitgegaan van een 'standaard' grondsoort voor geheel Europa. Daarmee bleef de complexiteit en het volume van de ontwikkeling nog redelijk binnen de perken. Het gehele systeem werd ontwikkeld binnen een combinatie van Fortran-programmatuur en Arc/Info-software. De door de Fortran-programma's gegenereerde gegevens worden opgeslagen in een uitgebreid stelsel van losse datafiles.

In deze voorfase is sterk gebruik gemaakt van een iteratieve ontwikkelaanpak, waarbij afwisselend het accent heeft gelegen op algoritme-ontwikkeling, met name voor ontbrekende weergegevens, gewasgroei-model-ontwikkeling en programmering. Deze fase is uitgevoerd met een relatief groot en sterk wisselend multidisciplinair team van verschillende onderzoeksinstituten. De inspanning van deze eerste fase omvat ca. 70% van de totaal bestede ontwikkelingsinspanning. Een eerste versie van het systeem is met de gegeven functionaliteit door het JRC als produktieversie gebruikt.

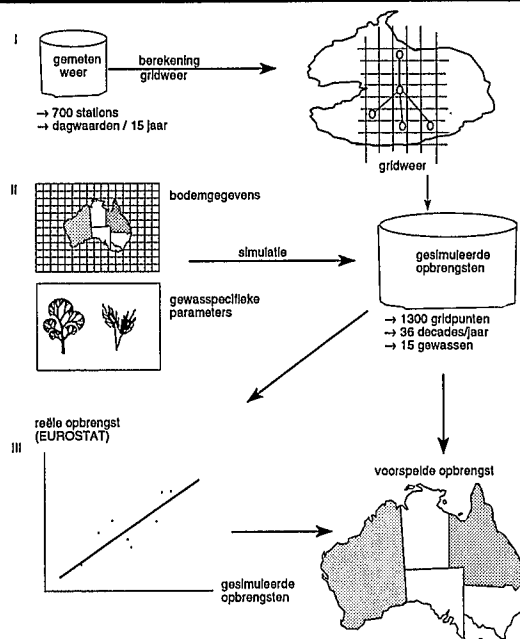
Redesign

De afspraken omtrent opleveringstijdstip van het gehele systeem waren aanleiding om in de tweede fase van het project de nodige randvoorwaarden en zekerheden in te bouwen. Zeker ook vanwege de reeds bestede inzet, de bestaande opzet van het systeem en documentatie, alsmede de nog aan te brengen complexiteit en volume van het systeem. Daardoor is een review-fase ingelast; in deze fase werd Agrimathica bij het project betrokken.

In eerste instantie is gekeken naar de bestaande sources, de onderlinge samenhang en consistentie. Geconstateerd werd dat vaak zeer verschillende functionaliteit ondergebracht werd binnen een en dezelfde Fortran-source; de onderliggende datafiles vertoonden veelal een sterke redundantie. Het volume aan uit te voeren berekeningen was mede debet aan deze aanpak; het systeem dient immers de berekeningen binnen redelijke tijd uit te voeren. Tevens bevatte de sources veel datamanipulatie-code. De iteratieve ontwikkelaanpak en de wisseling in het programmeursteam waren mede oorzaak van deze niet optimale situatie.

Aangegeven is dat het in deze fase van het project te overwegen was om alle datamanagement-functionaliteit in een database onder te brengen. Vanuit deze constatering is vervolgens besloten om boven de problematiek te gaan staan en een re-engineering fase op te starten. In deze fase zijn in een aantal sessies met inhoudelijk betrokkenen de systeemfuncties en onderliggende data opnieuw geanalyseerd en ondergebracht in een logische systeembeschrijving. Vervolgens is geko-

Figuur 1 - CGMS in vogelvlucht



zen voor een systeemarchitectuur gebaseerd op de volgende componenten:

- Een Oracle-database voor alle gegevens (weer, bodem, gewas), inclusief de resultaten van simulatieberekeningen;
- Oracle-schermen en -rapporten voor onderhoud van gegevens en relevante parameters;
- Arc/Info-database voor aanleveren van ruimtelijke informatie, alsmede het presenteren van resultaten op kaarten. De Arc/Info-database bevat geografische basisbestanden met bodems, landen, provincies, kustlijnen, hoogteligging en grid;
- Fortran-modules voor complexe, reken-intensieve taken.

Zie figuur 2 voor een illustratie van de systeemarchitectuur.

Op grond van deze keuzes is een projectplan opgesteld voor de tweede fase. Hierbij werd Agrimathica betrokken voor het engineering-gedeelte, c.q. de inbreng van de in de inleiding verklaarde PRODUCT-elementen. Een strakke fasering is aangebracht met duidelijke mijlpalen en commitment van alle partijen.

Omtrent de functionaliteit zijn met het JRC vaste afspraken gemaakt, die uitgangspunt vormden voor het op te leveren produkt. In deze tweede fase is zowel het (her-)ontwerp, de feitelijke ontwikkeling en documentatie uitgevoerd. De doorlooptijd voor deze tweede fase was ca. 9 maanden. De tweede fase van het project kon zo succesvol worden aangepakt door het vele inhoudelijke voorwerk dat was verricht. Algoritmen en modellen waren uitgekristalliseerd. De systeemeisen waren daardoor snel boven water. Even zo belangrijk is de flexibiliteit van de ontwikkelaars uit de eerste fase. Deze gingen mee in een andere (onbekende) systeemarchitectuur en technologie, en boden een onvoorwaardelijke steun aan de gekozen aanpak.

Oracle en CASE in een simulatie-toepassing

In dit project is gebruik gemaakt van Oracle*CASE. Alle 'common-data' is geana-

lyseerd en opgenomen in CASE*Dictionary. Gebleken is dat er meer 'common' is dan verondersteld. Tijdens de analyse zijn de simulatieroutines als het ware opengetrokken en ontdaan van alle datamanipulatiefuncties en gerelateerde data. Hierdoor kwam snel de onderliggende gegevensstructuur boven water. Tevens ontstond zo een goede scheiding tussen de diverse systeemfuncties; deze zijn vervolgens ondergebracht in afzonderlijke modules. CASE*Designer is gebruikt voor het vastleggen en definiëren van deze zaken in CASE*Dictionary.

Het gebruik van CASE vereist een systematische analyse, opsplitsing en beschrijving van het systeem. De gebruikte methodologie en schematechnieken, gebaseerd op IEM (Information Engineering Methodology), bleken voor de onderzoekers zeer behulpzaam bij de analyse. CASE vormt daarmee de basis voor de technische documentatie, die met reporting software grotendeels automatisch op maat wordt gegenereerd. CASE vormt daarmee de ruggraat voor het onderhoud van het systeem.

Een tweede punt waarvoor CASE is ingezet, is het automatisch genereren van invoerschermen en rapporten. Het systeem kenmerkt zich door relatief eenvoudige datamanipulatiefuncties, die daardoor goed met CASE*Generator te genereren waren. Voor deze modules is hierdoor de ontwikkeltijd aanzienlijk verkort.

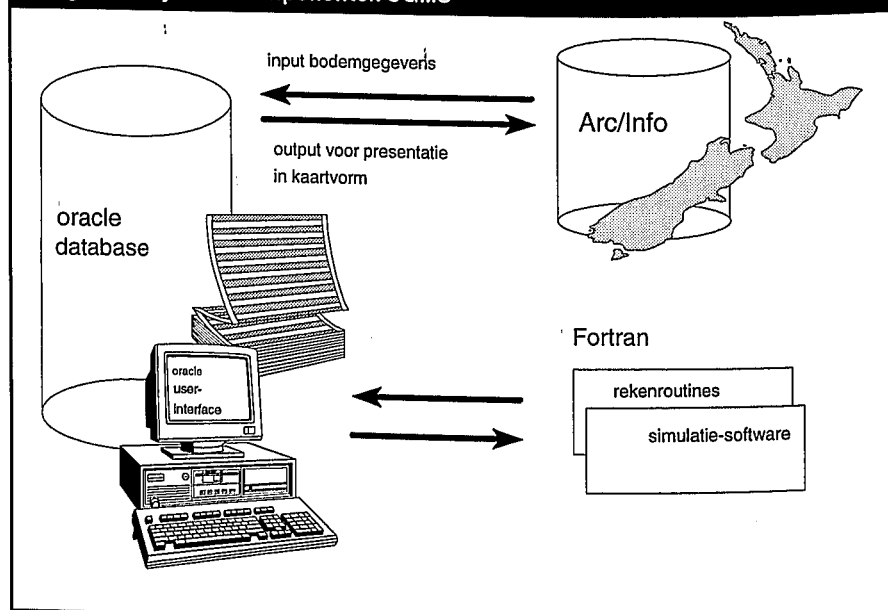
Vervanging van Fortran door Oracle

Aangegeven is dat veel van de Fortran-source code is vervangen door standaard Oracle-onderhoudroutines, PL/SQL-routines en embedded-SQL. De meeste I/O-operaties als filehandling, lees en schrijfopdrachten, inclusief sorteeralgoritmen en aggregatieprocedures konden worden vervangen. Daarmee kon de bestaande code tot ca. 40% worden gereduceerd. Voordeel van de aanpak is ook hier de onderhoudbaarheid van het systeem.

Performance en portabiliteit

De omvang van de database is ca. 5 Giga-byte. Dit legt uiteraard een zwaar beslag op de performance. Van relationele databasesystemen wordt vaak aangegeven dat performance een knelpunt is. In dit project zijn hiermee positieve ervaringen opgedaan. Daarvoor was wel enige database tuning nodig, omdat CASE uit zichzelf bij deze omvang geen optimale database genereert. Met de nodige inspanning konden sommige procedures met een gemeten 'performance' in termen van dagen, worden teruggebracht tot uren, hetgeen voor dit type toepassing van deze omvang zeer acceptabel is. Het opvoeren van performance is met name een kwestie van professionaliteit van de ontwikkelaar met een goede kennis van de hulpmiddelen. Het

Figuur 2 - Systeemcomponenten CGMS



bleek niet nodig om allerlei ondoorzichtige 'trucs' toe te passen.

De systeemontwikkeling heeft plaats gevonden op een VAX/VMS-platform. Het eindprodukt draait bij de eindgebruiker op een SUN-workstation onder UNIX; tijdens de ontwikkelfase is daartoe de code steeds 'geport' en vervolgens getest op de SUN. De sources bleken zeer portable, mede dankzij de stringente richtlijnen om tot een portable code te komen.

Tools en technologie voor hoge kwaliteit

Ook in dit project is duidelijk gebleken dat (re-)engineering niet alleen methodische en technische aspecten omvat. Het blijkt dat de uiteindelijke kwaliteit van software wordt bepaald door een groot aantal onderliggende factoren. Belangrijk is om vooraf het gehele ontwikkeltraject te beschouwen, niet alleen in grote lijnen, maar ook een vertaling te maken naar daadwerkelijk planbare werkpakketten. Gebleken is dat ook in een technisch-wetenschappelijke omgeving zaken te plannen zijn. Daarbij is belangrijk om tevoren zaken qua doelstelling en functionaliteit af te bakenen. Een gedegen ontwikkelaanpak (binnen dit project Information Engineering Methodology) helpt sterk bij het structureren van het project (figuur 3).

Een goede betrokkenheid van inhoudelijk deskundigen en gebruikers is een tweede belangrijke factor. Die betrokkenheid dient ook al in de allereerste fase te worden meegenomen. Over en weer dienen verwachtingen helder te zijn; impliciete aannamen van ontwikkelaars zijn uit den boze. Uiteindelijk bepaalt immers de gebruiker van het systeem de kwaliteit. De geboden professionaliteit in systeemontwikkeling blijkt een zeer grote rol te spelen bij de uiteindelijke, door de gebruikers ervaren kwaliteit.

Tools en technology staan in eerdergenoemd PRODUCT-rijtje achteraan. Goed

gereedschap en dito bouwstenen zijn evenwel een voorwaarde voor een goed eindprodukt. Professionaliteit van degenen die daarvan gebruik moeten maken is evenzeer van belang, zo niet belangrijker. Naarmate tools geavanceerder zijn, vereist dit een andere ontwikkeldiscipline en kennis. De toepassing van CASE-tools is immers niet triviaal. Het blijkt dat voor het gebruik van CASE inzicht nodig is van het gehele engineering-proces; het weten wat op welk abstractieniveau vereist is om tot een goed eindprodukt te komen. De aangedragen voorwaarden voor een succesvolle ontwikkeling hangen nauw samen met het idee wat kwaliteit inhoudt en hoe kwaliteit beheerst kan worden.

'Partner in Project'-concept

Dit project, waar reverse engineering een groot gedeelte van in beslag nam, is een voorbeeld van een aanpak zoals het DLO-Staring Centrum dit voorstaat voor het leveren van software aan derden. Systeemontwikkeling wordt aangepakt als een volwassen engineering-traject, waarbij alle aspecten van engineering goed ingevuld worden. Dit hoeft niet te conflicteren met de kwaliteit van de inbreng van het onderzoek. Door een juiste aanpak wordt de onderzoekruimte daardoor zelfs vergroot. Een systematische aanpak met gebruik van goede methoden maakt de projectrisico's in een vroegtijdig stadium zichtbaar. Het ontwikkelproces en de kwaliteit worden beheersbaar. De inschakeling van een IT-

partner als Agrimathica past daarbinnen. Er ontstaat een vruchtbare samenwerking tussen research engineers met goede kennis van de materie en modellen en professionele software engineers. Agrimathica hanteert bij deze aanpak de formule van partner of subcontractor. Het commitment dat daaruit voortvloeit heeft een groot effect naar de uiteindelijke eindgebruiker, hetgeen sterk bijdraagt in de kwaliteit en de continuïteit voor de gezamenlijke opdrachtgever.

Tot slot

In dit artikel is een praktijkcase aan de orde gekomen, waaruit duidelijk is dat software engineering en technisch-wetenschappelijk onderzoek kunnen samengaan. De aanpak, methoden en technologie uit de hoek van software engineering kunnen goed worden geïntegreerd binnen vergelijkbare projecten. Vanwege het thema van dit Agro-Informatica-nummer zijn de inzet van 4GL en CASE en de bijkomende voordelen besproken. Duidelijk is echter dat evenzoveel winst kan worden behaald uit de andere kwaliteitsbepalende elementen, zoals professionaliteit, projectorganisatie en gebruikersparticipatie.

De formule van een nauwe samenwerking tussen SC-DLO en Agrimathica op het gebied van software engineering voor kennisintensieve (agrarische) omgevingen heeft geresulteerd in een wetenschappelijk goed onderbouwd informatiesysteem met een hoge PRODUCTQUALITY.

Figuur 3 - Elementen voor kwaliteit in software-ontwikkeling

